

Spatial release from masking in reverberation for school-age children

Z. Ellen Peng,^{a),b),c)} Florian Pausch, and Janina Fels

Institute for Hearing Technology and Acoustics, RWTH Aachen University, Kopernikusstrasse 5, 52074 Aachen, Germany

ABSTRACT:

Understanding speech in noisy environments, such as classrooms, is a challenge for children. When a spatial separation is introduced between the target and masker, as compared to when both are co-located, children demonstrate intelligibility improvement of the target speech. Such intelligibility improvement is known as spatial release from masking (SRM). In most reverberant environments, binaural cues associated with the spatial separation are distorted; the extent to which such distortion will affect children's SRM is unknown. Two virtual acoustic environments with reverberation times between 0.4 s and 1.1 s were compared. SRM was measured using a spatial separation with symmetrically displaced maskers to maximize access to binaural cues. The role of informational masking in modulating SRM was investigated through voice similarity between the target and masker. Results showed that, contradictory to previous developmental findings on free-field SRM, children's SRM in reverberation has not yet reached maturity in the 7–12 years age range. When reducing reverberation, an SRM improvement was seen in adults but not in children. Our findings suggest that, even though school-age children have access to binaural cues that are distorted in reverberation, they demonstrate immature use of such cues for speech-in-noise perception, even in mild reverberation. © 2021 Acoustical Society of America. <https://doi.org/10.1121/10.0006752>

(Received 15 March 2021; revised 29 September 2021; accepted 29 September 2021; published online 2 November 2021)

[Editor: Emily Buss]

Pages: 3263–3274

I. INTRODUCTION

When learning in classrooms, children are constantly challenged by complex auditory environments. Attending to a target talker, amongst other competing talkers, is indisputably the most challenging listening situation for children (Buss *et al.*, 2017). As young as 4–5 years of age, children already demonstrate the ability to use auditory spatial cues for enhancing speech perception, specifically by receiving an intelligibility benefit when a target is spatially separated from the masker, as compared to when the two sound sources are co-located (Litovsky, 2005). This phenomenon is known as spatial release from masking (SRM), which is the result of gaining access to head shadow and binaural cues, namely, interaural time and level differences (ITD/ILD), during speech perception in the presence of interferers. When the interferer is similar to the speech target with high informational masking, such as speech babble rather than broadband noise, adults and children demonstrate larger SRM by better use of spatial cues in the absence of other cues for unmasking (Corbin *et al.*, 2017; Hawley *et al.*, 2004; Johnstone and Litovsky, 2006). Depending on the speech materials, some studies suggested that children show adult-like SRM in early childhood around 4–5 years of age (Ching *et al.*, 2011; Misurelli and Litovsky, 2012, 2015),

while others showed a more protracted trajectory into mid-childhood of 9–10 years (Cameron *et al.*, 2011; Van Deun *et al.*, 2010; Vaillancourt *et al.*, 2008; Yuen and Yuan, 2014). For most studies examining the developmental trajectory of SRM in children (Corbin *et al.*, 2017; Griffin *et al.*, 2019; and review by Yuen and Yuan, 2014), the experiments were conducted in sound booths that emulated free-field anechoic listening situations. Little is known about whether such findings generalize to indoor environments, which children encounter daily in realistic learning situations.

Unique to indoor spaces, reverberation is the result of sound energy reflected from interior surfaces that arrives at the listener with a slight delay and reduced energy after the initial direct sound (Kuttruff, 2009). It reduces the interaural coherence between the signals arriving at the two ears (Blauert, 1997; Blauert and Lindemann, 1986a). In the precedence effect literature (Litovsky *et al.*, 1999; Brown, Stecker, and Tollin 2015 for reviews), the phenomenon that adult listeners can accurately localize sounds in reverberant rooms shows that human listeners have substantial tolerance for reverberation (Goupell *et al.*, 2012; Hartmann, 1983; Rakerd *et al.*, 2010; Rakerd and Hartmann, 1985). Yet, human listeners are highly sensitive when asked to detect interaural incoherence (Goupell and Hartmann, 2006), which is a hallmark of distorted binaural cues from reverberation. Many have suggested that reduced interaural coherence leads to the perceptual consequence of a widened auditory source width (Blauert and Lindemann, 1986b, 1986a; Hidaka *et al.*, 1995; Johnson and Lee, 2019;

^{a)}Electronic mail: ellen.peng@boystown.org, ORCID: 0000-0002-3487-2144.

^{b)}Also at: Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705, USA.

^{c)}Currently at: Boys Town National Research Hospital, 555 N. 30th Street, Omaha, NE 68131, USA.

Robinson *et al.*, 2013; Whitmer *et al.*, 2013, 2014). In short, even though adults can generally localize sounds at various locations in reverberant rooms, the auditory spatial cues associated with the spatial separation between source locations are less salient than in the free-field environment.

Others have probed listeners' use of binaural cues distorted from reverberation through SRM, which is quantified as the speech reception threshold (SRT) improvement (in dB) when target and maskers are spatially separated versus when the sound sources are co-located. Freyman *et al.* (1999) found that adult listeners demonstrated smaller SRMs when an artificial, single reflection was introduced. Kidd and colleagues (2005; Marrone *et al.*, 2008) measured SRM in various reverberant environments by altering the interior materials in a sound booth to increase the reverberation time (RT), which subsequently reduced the interaural coherence. They showed that young adult listeners' SRM substantially reduced by ~ 4 dB when RT slightly increased from 0.06 to 0.25 s. Other studies showed a similarly deleterious effect of reverberation on SRM by systematically changing reverberation, either through room recordings or acoustic simulations that were binaurally reproduced to the listeners in virtual reality (Beutelmann and Brand, 2006; Deroche *et al.*, 2017; Muñoz *et al.*, 2019; Rennie *et al.*, 2011; Rennie and Kidd, 2018; Ruggles *et al.*, 2011). For a spatial configuration with a frontally positioned target, Rennie and Kidd (2018) outlined two aspects of interaural incoherence due to reverberation for maskers displaced at off-center locations. First, the ILD is smaller because reflected sounds compensated for the attenuated signal in the contralateral ear (Shinn-Cunningham *et al.*, 2005), which also resulted in a smaller head shadow effect. Second, the ITD may be less salient because the overlapping reflected sounds share highly similar spectro-temporal characteristics with the direct sound, making the extraction of ongoing ITD through cross correlation generally much more difficult. Further, there is evidence to suggest that the reduced salience of binaural cues reduced auditory attention and impaired stream segregation (Oberem *et al.*, 2018; Ruggles *et al.*, 2011). Adding to the difficulty of using distorted ILD and ITD cues, reverberation also smears the temporal envelope by reducing amplitude modulation depth that is critical for speech perception (Deroche *et al.*, 2017; Houtgast and Steeneken, 1985; Nábelek *et al.*, 1989).

While the detrimental impact of reverberation on SRM has been studied extensively for adults, to the best of our knowledge, there is no published study on similar investigations with children. Work on auditory development in the precedence effect by Litovsky (1997) and Litovsky and Godar (2010) provided some useful hints on how children's spatial hearing may be affected when using distorted binaural cues. In a minimum audible angle (MAA) task to measure spatial acuity, 5-year-old (YO) children who had adult-like MAA in free-field conditions showed higher MAA thresholds than adults when a single reflection was introduced (Litovsky, 1997). In a sound localization task, also in

the presence of a single reflection, Litovsky and Godar (2010) compared echo thresholds and localization errors between 4 and 5 YO children and adults. The young children showed elevated echo thresholds by ~ 10 ms in identifying trials containing two sound sources. For trials identified with a single sound source, they also had higher root mean square errors for localization. Both studies implied that for children, at least until the age of 5, the ability to detect interaural incoherence has not yet reached maturity.

The main goal of the present study is to examine the impact of reverberation on normal hearing children's use of binaural cues for speech understanding. Along with the refinement of using auditory spatial cues (Corbin *et al.*, 2017; see Yuen and Yuan, 2014 for review), during the critical developmental window in the first 10–15 years in life, children are also developing other auditory skills to aid speech-in-noise segregation, such as spectral/temporal resolution and glimpsing (Buss *et al.*, 2017; Leibold and Buss, 2019; Leibold and Neff, 2007, 2011). The development of children's ability to use binaural cues distorted by reverberation for SRM in reverberation during this age range, as compared with adults, is the main focus of this investigation.

Notably, the detrimental effect of reverberation on children's speech perception has been studied extensively. Past literature has repeatedly demonstrated the deleterious impact of reverberation on children's speech perception, including phoneme identification (Johnson, 2000; Klatte *et al.*, 2010; Nábelek and Robinson, 1982; Yang and Bradley, 2009), sentence recognition (McCreery *et al.*, 2019; Neuman *et al.*, 2010, 2012; Wróblewski *et al.*, 2012), and pragmatic comprehension (Valente *et al.*, 2012). As reverberation and noise co-occur in most indoor spaces, studies have shown the additive effect of reverberation in further impairing speech-in-noise processing (Neuman *et al.*, 2010; Peng and Wang, 2016; Prodi *et al.*, 2013). For children, different types of maskers have a differential impact on speech understanding (Johnstone and Litovsky, 2006; Wróblewski *et al.*, 2012). The most challenging type of masker is the two-talker babble (Buss *et al.*, 2017; Leibold and Buss, 2013). For children, their ability to use spectro-temporal cues to stream segregate target from two-talker babbles does not reach maturity until late adolescence (Buss *et al.*, 2017; Leibold and Buss, 2019). Besides factors in sensory processing, there is also evidence to suggest that poorer speech-in-noise performance is in part due to elevated internal noise and/or immature executive functions (e.g., non-sensory processing) in the auditory domain (Cabrera *et al.*, 2019; Maccutcheon *et al.*, 2019; McCreery *et al.*, 2019).

To examine how realistic reverberant environments affect children's use of binaural cues for speech understanding, we looked to state-of-the-art virtual reality technologies to measure behavioral outcomes. A system with a four-channel loudspeaker array was recently built to deliver virtual acoustic environments (VAEs) for conducting psychoacoustic experiments in a sound booth (Pausch *et al.*, 2018). This system can reproduce spatial sounds in VAE

through a cross talk cancellation algorithm (Masiero and Vorländer, 2014) with high perceptual fidelity similar to spatial perception outcomes in a free-field setup (Pausch and Fels, 2020). The audio reproduction system is supported by a back-end engine (Wefers, 2015) to create real-time VAE that simulated reverberant auditory environments based on acoustic models of typical classrooms (Schröder and Vorländer, 2011). The application in this investigation provided an unprecedented opportunity to measure behavioral psychophysics of children in a virtual auditory environment that mimicked real-world listening, using a laboratory-controlled experimental design. Specifically, it allowed a within-subject design protocol by providing acute exposure to different reverberant environments. All children completing the study were able to finish all testing (eight conditions) during a single test session of <2 h. This highly controlled experimental design allowed for rigorous behavioral comparisons during data analysis, leading to improved validity in our findings to infer behavioral outcomes for children in the real world.

In this investigation, by utilizing acoustic simulation techniques, we aimed to answer the following research questions:

- (1) What is the effect of reverberation on children's SRT? Specifically, do children demonstrate any improvement in SRT by reducing RT from 1.1 to 0.4 s?
- (2) What is the effect of reverberation on children's SRM? Specifically, do children demonstrate any improvement in SRM by reducing RT from 1.1 to 0.4 s?
- (3) What is the role of masker similarity with the target (i.e., same- versus different-voice) in moderating the impact of reverberation on SRM?

We hypothesized that increasing RT would lead to a larger elevation in SRT and reduction in SRM for children between 7 and 12 years old than for adults due to immature processing of distorted binaural cues. We chose the age range of 7–12 years, as it encompasses the critical period of auditory development that may reveal important developmental milestones of SRM (Cameron *et al.*, 2011). These children were split into two groups of 7–9 and 10–12 YO, as previous work has suggested 9–10 years of age may be a maturation landmark for masked speech recognition (see review by Leibold and Buss, 2019). In this investigation, we manipulated the two-talker babble masker similarity by varying between the same versus different (but same-sex) talker as the target, a condition that has never been investigated previously in reverberation. We hypothesized that, by completely removing the voice cues, both adults and children may rely more on spatial cues for unmasking, leading to a larger increase in SRM with reducing RT as compared to such change when the maskers had different voices than the target. Alternatively, if auditory spatial attention is impaired by reverberation, we predicted a smaller change in SRM between the two RTs in the same-voice condition than in the different-voice condition.

II. MATERIALS AND METHODS

A. Participants

Eighteen children and 16 adults completed the study. The children were split into two groups based on age. The younger children group had nine children aged between 7 and 9 years old [7 years, 7 months to 9 years, 4 months ($M = 8.9$ years)], and the older children group had nine children aged between 10 and 12 YO [10 years, 1 month to 12 years, 7 months ($M = 11.5$ years)]. All children had normal hearing based on annual hearing screen performed at the school, as well as parental reports on the day of testing. None of the children had language impairment or were on medication during their participation in the study. A control group of 16 adults ($M = 25$ years) also participated in the study. All adults passed a hearing screen with hearing thresholds ≤ 20 dB hearing level in both ears from 125 to 8000 Hz before participating in the experiment. All child and adult participants are native German speakers. The experimental procedure was approved by the university hospital ethical commission at RWTH Aachen University (EK 188/15).

B. Reverberation Simulation

To create reverberant environments, we modeled a typical rectangular classroom of 244 m^3 ($11.8 \times 7.6 \text{ m}$ in area, and 3 m height) in the Room Acoustics Virtual Environment (RAVEN) software, which was developed at the Institute of Technical Acoustics (Pelzer *et al.*, 2014; Schröder and Vorländer, 2011). Simple furniture, such as wooden desks, was added to the model to improve realism. A listener location was defined in the front row of the virtual classroom with a target talker location in front (0° azimuth) at 2 m away as the typical teacher location. Two additional sound sources were positioned at equal distance symmetrically away from the listener on two sides at $\pm 90^\circ$ azimuth as the maskers in conditions with a 90° spatial separation between the target. By changing the interior surface absorption and scattering coefficients, two RT conditions were simulated at 0.4 s and 1.1 s, averaged across octave frequency bands from 500 to 2000 Hz. The choice of reverberant environments was based on the maximum RT of 0.6 s recommended by the American National Standards Institute (ANSI) (Acoustical Society of America, 2002) for critical learning spaces for K–12 education. We chose one RT that was slightly better than the ANSI recommendation at 0.4 s, and the other at 1.1 s that exceeded the recommendation but was common in large lecture halls. Figure 1 shows the simulated RT in T_{30} across octave frequency bands for the two RT conditions. The binaural room impulse responses (BRIRs) were simulated using a set of head-related transfer functions (HRTFs) that were scaled from a standard manikin (see more details in Sec. II C). Reverberation was modeled using a hybrid algorithm to simulate early reflections (i.e., up to second order) through an image source method and late reflections through ray-tracing (Schimmel *et al.*, 2009). All BRIRs were pre-generated to allow an individualization

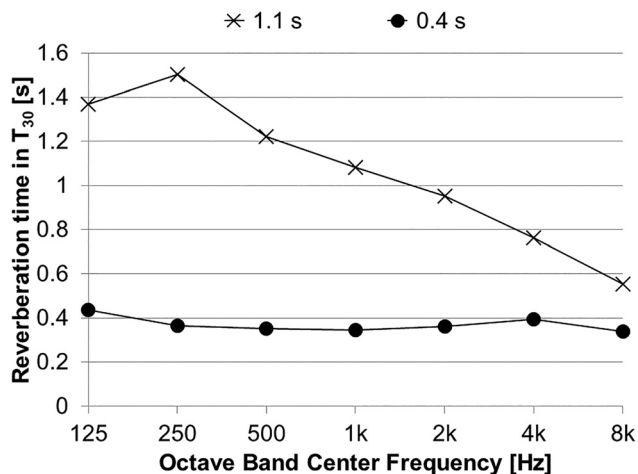


FIG. 1. Reverberation time in T_{30} as a function of octave band center frequencies from 125 to 8000 Hz for the two simulated reverberation conditions.

procedure described in Sec. II C and convolved with the speech recordings in real-time during stimulus presentation.

C. Binaural Reproduction

Children's anthropometrics, including head size, change as they grow (Clifton *et al.*, 1988; Fels *et al.*, 2004). Rendering virtual auditory scenes with a set of HRTFs from the standard manikin that was modeled after a median sized adult, lacked the ecological validity in simulated BRIRs experienced by children in real-world listening. However, the limited experimental time with children did not allow for recording individual HRTFs. Hence, we used an individualization procedure to select a set of pre-generated HRTFs that best fit the child's head size. The HRTFs recorded from a standard manikin (Schmidt, 1995) in 3° resolution on the horizontal plane were scaled both in interaural time difference cues (Bomhardt *et al.*, 2016, 2018; Bomhardt and Fels, 2014) and spectral cues (Middlebrooks, 1999). The pre-generated database was created based on three head sizes: the 15th, 50th, and 85th percentile of typical child anthropometric measures of depth, height, and width (Fels, 2008; Fels *et al.*, 2004; Fels and Vorländer, 2009). Together, 27 sets of scaled HRTFs were available for selection to create the individualized virtual auditory scenes. At the beginning of the experimental session, each child's head size (i.e., depth, height, and width) was measured; one of the 27 HRTF sets was then selected, which shared the closest dimensions with the child's head size. This individualization procedure best approximated the realistic binaural cues delivered in the virtual auditory environment for use with children in the laboratory setting. For adult listeners, the standard BRIRs with HRTFs from the standard manikin were used.

All testing was done in a sound-insulated booth (2.1 m by 2.1 m in area and 2 m in height) with a volume of $\sim 9 \text{ m}^3$. A four-channel loudspeaker array (KH-120A, Georg Neumann GmbH, Berlin, Germany) with cross talk

cancellation was used to provide binaural reproduction of the virtual classroom scene (Pausch *et al.*, 2018). The loudspeakers were positioned at the four corners and approximately at ear height for the children. An audio interface (RME Fireface UC, Audio AG, Haimhausen, Germany) was used to present stimuli from custom MATLAB software (Mathworks, Natick, MA). To calibrate in VAEs, we used speech-shaped noise that matched the long-term averaged spectrum of the target and two maskers, and normalized the root mean square pressure levels between the two RT conditions. Even though speech levels are typically higher in more reverberant environments than environments with shorter RTs, we kept the overall speech levels the same across conditions to remove potential confounds due to increased audibility. In both RT conditions, the SNRs were maintained. During the experiment, listeners were seated at the center of the sound booth at approximately 110 cm away from the loudspeakers. Listeners were told that they could move their heads during the experiment if desired. Head position and orientation were tracked by four optical motion capture cameras (Flex 13, NaturalPoint, Inc. DBA OptiTrack, Corvallis, Oregon, United States). The system continuously updated the direct sound portion in the BRIRs (see Pausch *et al.*, 2018; Table II, configuration A) to correct for offsets away from the midline in the horizontal plane, so that the target sound source always appeared at $\sim 0^\circ$ ($\pm 1.5^\circ$) azimuth relative to the listener in the virtual scene. The dynamic end-to-end system latency was estimated to be no longer than 84 ms, providing realistic and responsive dynamic binaural reproduction of the virtual acoustic scenes (Pausch *et al.*, 2018).

D. Experimental Design

The experimental paradigm used in the present study was inspired mainly by the Listening in Spatialized Noise Test-Sentences (LiSN-S) developed by Cameron and Dillon (2007), who assessed SRT and SRM by children in free-field environments using a sentence recognition task with two-talker maskers that were either the same voice of the target or different voices (but same-sex) from the target. We adapted the paradigm by utilizing German speech materials and implemented testing in VAEs with simulated room acoustics.

1. Speech materials

A subset of age-appropriate five-word sentences from the German HSM Sentence Test (Hochmair-Desoyer *et al.*, 1997) was chosen as the target speech. An example HSM target sentence is "Bitte zeig mir Deinen Ring." Masker speech was continuous discourse from eight Grimm stories in German that were generally unfamiliar to children. Three female native German speakers were recruited to record these materials in an anechoic chamber and were instructed to speak at conversational speed. Speech materials were recorded using a digital recorder (ZOOM H6, Sound Service GmbH, Berlin, Germany) and a condenser microphone (TLM

170, Georg Neumann GmbH, Berlin, Germany) set to a cardioid directivity pattern at 44.1 kHz sampling frequency.

The speech characteristics of the three female talkers are listed in Table I. Talker A was chosen as the target speaker and recorded all HSM sentences and eight stories; two other talkers recorded only the masker stories as their voices shared similar but slightly lower fundamental frequencies. Acoustic analysis was done in Praat [Version 6.0.18; (Boersma and Weenink, 2021)] on a ~60 s speech sample from the same Grimm stories for all three talkers. A 10 s Gaussian-shape analysis window with 25% overlaps was set up for measuring the first and second formants, which were reported as the averaged values in Table I. As seen in the Table, the two masker talkers had slightly lower fundamental frequencies than the target talker; all three talkers shared similar speech rates.

2. Test conditions

While being immersed in the virtual acoustic environment provided through the dynamic binaural reproduction, each listener was tested for eight conditions consisting of 2 RTs (0.4 s versus 1.1 s) \times 2 voice conditions (target-masker same- versus different-voice) \times 2 spatial conditions (target-masker co-located versus spatially separated at 90° symmetrically).

In the different-voice conditions, with voice cues available for stream segregation, two female talker voices (i.e., Talkers B and C) that were different from the target talker (Talker A) were used for the maskers. In the same-voice conditions, we removed voice cues by presenting the Grimm stories recorded by Talker A, who share the same voice as the target talker. Although contextual cues still existed in the same-voice condition, important voice cues, such as mean F0 and speech rate, were removed by using the same voice between the target and maskers. For the spatial conditions, the two-talker maskers varied in the azimuthal position either together at 0° (co-located with the target) or symmetrically displaced at $\pm 90^\circ$ (spatially separated from the target) in VAE. By using the configuration of symmetrically displaced maskers, we minimized access to head shadow (as opposed to the asymmetrical configuration) and maximized listeners' access to utilize ITD and ILD cues for SRM (Misurelli and Litovsky, 2012).

3. Speech intelligibility measurement

Speech reception thresholds (SRT) were measured adaptively using a one-up-one-down adaptive procedure (Levitt, 1971) to obtain 50% speech intelligibility. Masker

speech level was fixed at 55 dB SPL, and its onset always preceded the first target sentence by 3 s and ended at the same time. For each test condition, the target speech was initially presented at 70 dB SPL with an initial +15 dB SNR and followed a 4 dB step size to either reduce or increase SPL depending on the listener's responses until the first reversal, after which the step size changed to 2 dB. The maximum allowable target level was set at 80 dB SPL, to ensure that all stimulus presentations were safe and comfortable for children but was never reached by any child or adult during testing.

To score for accuracy, an experimenter sitting outside the sound booth listened to the verbal response and chose the keywords repeated correctly. A sentence with three or more keywords repeated correctly resulted in reducing target speech level in the corresponding step size for the next test trial. A test condition was terminated when the listener reached seven reversals. SRTs were calculated by fitting a sigmoid function to all the trial data in each test condition and extracting the SNR at 50% accuracy, using methods described by Frund *et al.* (2011) and the accompanying MATLAB “psignifit” toolbox (version 3). Supplementary materials include illustrations of an example curve fit¹ for SRT for one child and individual variability of estimating SRT as the 95% confidence intervals¹ that are similar among the listener groups across test conditions.

4. Procedure

Parental consent was obtained for each child before they participated in the experiment. All adult participants provided written consent. After measuring the head dimensions, listeners were seated in the listening booth. They were instructed to verbally repeat each target sentence and ignore the masker speech during the task. A familiarization phase was given using the HSM sentences with three keywords to ensure that all listeners understood the task. Listeners passed the familiarization phase if they scored 100% in the first five sentences or until reaching a cumulative 80% correct. All listeners, both children and adults, succeeded in completing the initial familiarization phase. During the testing phase, the experimenter initiated each test trial once the child was ready and marked all keywords correctly repeated by the listener. We assigned the order of presentation of the eight test conditions following a Latin Square design for counter-balancing. Even though perfect counter-balancing was not reached, the order of condition presented as a single fixed effect predictor (with listeners as the random effect) did not significantly affect SRT ($p > 0.05$). In general, when averaged across all listeners, SRT improved by 0.3 dB from the first to the last experimental run, which suggested that any potential order effect was washed out. These children were able to complete the task in a single 2 h visit that also included informed consent/assent, verbal confirmation for hearing screen with parents; they were given frequent breaks when needed. Each child

TABLE I. Speech characteristics for the three female talkers used in this study.

	F0 [Hz]	F1 [Hz]	F2 [Hz]	Speech rate (syllables/s)
Target talker	213	619	1478	3.2
Masker talker 1	191	580	1630	3.2
Masker talker 2	198	562	1611	3.4

was compensated 10 € for the study. Adult participants did not receive payment for their participation.

III. RESULTS

Statistical analysis was performed using R (version 4.0.2) and R Studio (1.3.1073). As a general approach, a linear mixed-effects model was fitted to the dependent variable of either SRT or SRM. The initial full model included individual participants as the random effect and all main effects (e.g., age group, spatial cue, voice cue, and RT) and interactions as the fixed effects. We then applied a backward elimination procedure (“buildmer” package) using the likelihood-ratio tests for elimination criterion to reduce fixed effects that did not significantly contribute to the model prediction of the dependent variable. The final model is reported. For *post hoc* analysis of the age group main effect, we used pairwise comparison (i.e., Wilcoxon Rank Sum Test) with Bonferroni correction (Field *et al.*, 2012). For *post hoc* analysis on interactions, we chose one of the main effects (e.g., voice cue) and fitted additional linear mixed models on each level. In addition, we performed simple effects analysis to examine the effect of reverberation, by comparing performances (SRT or SRM) between two RTs in each age group through the Wilcoxon Signed Rank Test under individual test conditions. An *a priori* significance level at 0.05 was set for all statistical models.

A. Speech Reception Thresholds

Figure 2 illustrates the average SRT by RT for each age group in the specific spatial-voice condition. After dimension reduction, the maximal linear mixed-effects model suggested statistical significance for all main effects: age group [$F(2, 31) = 8.9$, $p = 0.00014$], RT [$F(1, 31) = 48.7$, $p < 0.0001$], spatial cue [$F(1, 31) = 599.2$, $p < 0.0001$], and voice cue [$F(1, 31) = 4.4$, $p = 0.037$]. *Post hoc* analysis on

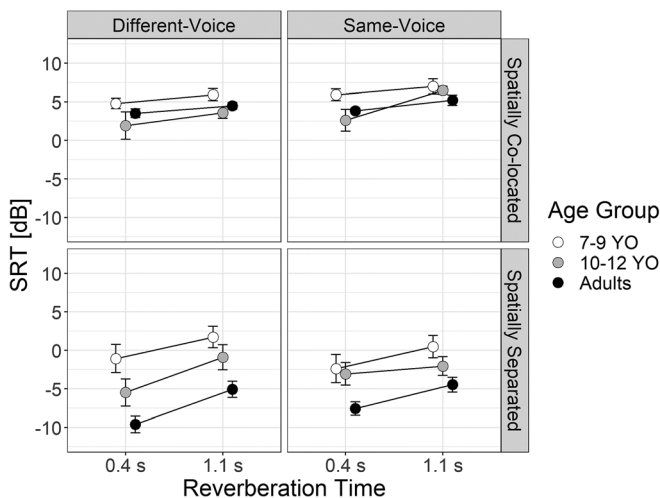


FIG. 2. Mean SRTs by reverberation time condition for each spatial-voice condition for three age groups: 7–9 YO (white circles), 10–12 YO (gray circles), and adults (black circles). Error bar represents one standard error of the mean (SEM).

the age group main effect suggested that 7–9 YO younger children as a group ($M = 2.8$ dB, $SD = 2.6$) had significantly elevated SRTs than 10–12 YO older children ($M = 0.4$ dB, $SD = 2.5$, $p = 0.027$) and adults ($M = -1.2$ dB, $SD = 2.2$, $p < 0.001$), respectively. The group average SRTs did not differ between 10 and 12 YO older children and adults. When averaged across other conditions, reducing RT from 1.1 s to 0.4 s resulted in an SRT improvement of 1.6 dB. By introducing a 90° spatial separation between the target and maskers, on average, listeners experienced an SRT improvement of 11.8 dB as compared to when target and maskers were co-located. The voice cue, by introducing two different female maskers as compared to both maskers having the same voice as the target, provided a small but statistically significant SRT improvement of 0.7 dB.

The final model also revealed two significant two-way interactions: age group \times spatial cue [$F(2, 31) = 21.5$, $p < 0.0001$] and RT \times spatial cue [$F(1, 31) = 6.0$, $p = 0.014$]. While there existed a large SRM of 11.8 dB on average as shown in the main effect of spatial cue, both significant interactions involved the spatial cue main effect, suggesting the magnitude of SRM was dependent on both age and RT. The effect of age and RT on SRM is further examined in Sec. III B.

The present study was the first attempt to examine the impact of reverberation under the presence of various cues across age groups. We hence performed additional follow-up simple effects analysis as a follow-up to examine how SRTs varied between 0.4 s versus 1.1 s RT in each test condition per age group. Table II listed the results from the Wilcoxon signed rank tests for the simple effects. In the different-voice conditions, both groups of children and adults showed similar trends. When the target and maskers were spatially co-located, SRT did not significantly improve with reducing RT. However, when the maskers were spatially separated by 90° from the target in VAE, all listeners demonstrated a significant improvement in SRT when RT reduced from 1.1 to 0.4 s. In the same-voice conditions, the effect of RT was more variable across listener groups. Adults and older children of 10–12 YO old demonstrated SRT improvement with reducing RT in both spatial conditions, but not the younger children of 7–9 YO. Neither group of children seemed to receive substantial improvement from lowering RT when the target and maskers were spatially separated.

B. Spatial Release from Masking

In the present study, SRM was defined as the intelligibility benefit due to the 90° target-masker spatial separation, as compared to when the target and maskers were co-located. We quantified SRM by taking the difference between SRTs measured in the spatially co-located and separated conditions. Four SRMs were calculated for each listener across the two voice cues \times 2 RTs test conditions using the equation below:

$$SRM = SRT_{\text{Spatially Co-located}} - SRT_{\text{Spatially Separated}}.$$

TABLE II. Mean and standard deviation (SD) of speech reception threshold (SRT) for each age group. Wilcoxon Signed Rank tests compared SRT measured from the 0.4 s versus 1.1 s reverberation time (RT) condition. Significant *p*-values are highlighted in bold.

Age group	Voice Cue	Spatial Cue	SRT (SD)		Wilcoxon Signed Rank Test
			RT = 0.4 s	RT = 1.1 s	
7–9 YO	Difference-Voice	Spatially co-located	4.8 (2.1)	5.9 (2.4)	<i>p</i> = 0.250
		Spatially separated	−1.1 (5.5)	1.7 (4.2)	<i>p</i> = 0.004
	Same-Voice	Spatially co-located	5.9 (2.3)	7.0 (2.9)	<i>p</i> = 0.500
		Spatially separated	−2.4 (5.5)	0.5 (4.3)	<i>p</i> = 0.200
10–12 YO	Difference-Voice	Spatially co-located	1.9 (5.3)	3.6 (2.2)	<i>p</i> = 0.360
		Spatially separated	−5.5 (5.3)	−0.9 (4.9)	<i>p</i> = 0.020
	Same-Voice	Spatially co-located	2.6 (4.2)	6.5 (1.5)	<i>p</i> = 0.020
		Spatially separated	−3.1 (4.4)	−2.0 (3.6)	<i>p</i> = 0.360
Adults	Difference-Voice	Spatially co-located	3.5 (2.2)	4.5 (1.7)	<i>p</i> = 0.140
		Spatially separated	−9.6 (4.2)	−5.1 (4.2)	<i>p</i> < 0.001
	Same-Voice	Spatially co-located	3.9 (1.5)	5.2 (2.7)	<i>p</i> = 0.039
		Spatially separated	−7.6 (3.5)	−4.5 (3.8)	<i>p</i> = 0.0023

Figure 3 shows the average SRM by RT for each age group, separately for the same versus different voices conditions in each panel. The linear mixed-effects model fitted to the SRM data revealed significant main effects for age group [$F(2,31) = 7.5$, $p = 0.00057$] and RT [$F(1, 31) = 8.2$, $p = 0.0042$]. *Post hoc* analysis on the age main effect suggested that both groups of children had significantly smaller SRM ($M = 6.2$ dB, $SD = 5.0$, $p < 0.0001$ for 7–9 YO; $M = 6.5$ dB, $SD = 5.7$, $p = 0.0001$) than adults ($M = 10.9$ dB, $SD = 4.0$). There was no significant difference in the average SRMs between the two groups of children ($p > 0.05$). When averaged across all other fixed effects, listeners demonstrated an improvement of 2.9 dB SRM when reducing RT from 1.1 to 0.4 s.

After backward elimination from the initial full model, the final model included voice cue even though its main effect was not significant ($p = 0.36$). The final model also revealed a significant two-way interaction between RT and voice cue [$F(1, 31) = 4.0$, $p = 0.046$]. For *post hoc* analysis, we fitted a linear mixed-effects model separately for each level of the voice cue fixed effect (same- vs different-voice

between the target and maskers). The same backward elimination procedure was applied for dimension reduction in the *post hoc* models. The *post hoc* model included age group and RT as fixed effects and individual listeners as the random effect. In the different-voice condition, both the age group ($p < 0.0001$) and RT ($p = 0.00014$) main effects were statistically significant in the *post hoc* linear mixed-effects model. Reducing RT generally provided a 2.9 dB SRM improvement as averaged across all age groups. In the same-voice condition, only the age group was included in the final *post hoc* model with significant main effect ($p = 0.020$). The RT fixed effect was excluded during the backward elimination procedure, and its main effect was non-significant ($p > 0.05$).

Table III shows the results of paired comparison of SRMs from 0.4 s and 1.1 s RTs within each age group. More specifically, the SRM improvement by reducing RT was only seen in adults but not in either group of children. For all three groups of listeners under 1.1 s RT, there was no significant difference in the average SRM between the two voice cue conditions.

Figure 4 illustrates the individual data of SRM measured under 0.4 s versus 1.1 s RTs. From the individual SRM, we identified one child in the 10–12 YO group who was tested in the same-voice condition at −10 dB SRM in 0.4 s RT, below 2 standard deviations (SD) of the group distribution; but the same child was tested within 1 SD in 1.1 s RT. By removing this child and re-fitting the *post hoc* linear mixed-effects model for the same-voice condition, RT main effect became statistically significant ($p = 0.0052$). When the target and maskers shared the same voice, lowering RT provided a significant SRM improvement but of smaller magnitude at 1.7 dB than in the different-voice condition. Removal of this child in the comparison shown in Table III (same-voice condition only) did not change the statistical outcome in the 10–12 YO children group.

The diagonal lines in Fig. 4 highlighted equal SRMs (± 2 dB) from both RTs. For individual data plotted in the area under the diagonal line, the listeners demonstrated

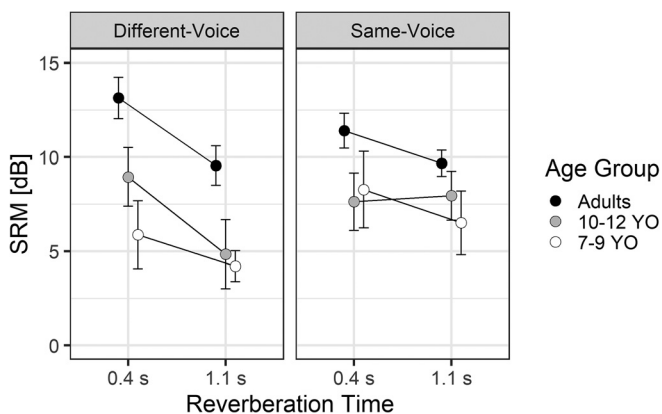


FIG. 3. SRM by reverberation time measured with target and maskers sharing the same voice (left panel) and different voices (right panel). Mean ± 1 SEM are plotted for three age groups: 7–9 YO (white circle), 10–12 YO (gray circle), and adults (black circle).

TABLE III. Mean and standard deviation (SD) of spatial release from masking (SRM) for each age group. Wilcoxon Signed Rank tests compared SRM measured from the 0.4 s versus 1.1 s reverberation time (RT) condition. Significant p-values are highlighted in bold.

Age group	Voice cue	SRM: Mean (SD)		Wilcoxon Signed Rank Test
		RT = 0.4 s	RT = 1.1 s	
7–9 YO	Difference-voice	5.9 (5.4)	4.2 (2.5)	p = 0.16
	Same-voice	8.3 (6.1)	6.5 (5.1)	p = 0.57
10–12 YO	Difference-voice	7.4 (6.2)	4.5 (4.9)	p = 0.16
	Same-voice	5.7 (7.2)	8.6 (3.9)	p = 0.65
Adults	Difference-voice	13.1 (4.4)	9.5 (4.2)	p = 0.0076
	Same-voice	11.4 (3.7)	9.7 (2.8)	p = 0.044

larger SRM from 0.4 s RT than 1.1 s RT; whereas, if the individual data were located above the equal SRM area, listeners received larger SRM from 1.1 s RT than 0.4 s RT. We also observed that group averaged SRMs plotted with crossed error bars in red were further away from the equal SRM area for adults, while the crossed error bars for both groups of children were mainly overlapping with the equal SRM areas. In general, children's data were more spread out, showing larger individual variability than adults. There were consistently more adults than children with individual SRMs located in the area below the equal SRM line, suggesting reduced SRM with longer RT.

IV. DISCUSSION

In this study, we investigated 7–12 YO children's speech-in-noise intelligibility and SRM in simulated

reverberant indoor environments of 0.4 s and 1.1 s RT that mimicked their typical learning environments. For the SRM measure, we used a spatial configuration in VAE which specifically examined children's use of binaural cues for SRM in reverberant environments that are known to reduce the salience of such cues. The novelty of using VAE and binaural reproduction in a sound booth enabled a robust experimental design of introducing acute exposure of reverberation within the same test session.

Compared to adults, children are more vulnerable to informational masking presented by maskers that share more speech-like characteristics with the target, such as speech babbles than modulated noise (Wightman *et al.*, 2010). When tested in free-field, children demonstrate a more protracted developmental trajectory into adolescence when perceiving speech in noise with high informational masking, such as when the target speech was masked by speech babble (i.e., two-talker maskers) than stationary or modulated speech-shaped noise (review by Leibold and Buss, 2019). How realistic reverberation influences masked speech perception with high informational masking is less well known with studies mostly focused on using speech-shaped noise (McCreery *et al.*, 2019; Wróblewski *et al.*, 2012; Yang and Bradley, 2009). Our investigation, with findings on adult-like SRTs in reverberation among children between 10 and 12 YO, did not detect the late maturation of speech-in-speech perception (i.e., two-talker masker) that was otherwise demonstrated in free-field (Buss *et al.*, 2017). In the present study, even though reducing RT from 1.1 to 0.4 s on average provided a 1.6 dB improvement in SRT, the effect varied between the two spatial conditions. In the different-voice conditions across all three age groups, such intelligibility benefit from reducing reverberation was only observed in the spatially separated condition, when binaural cues were available, but not in the co-located condition (see Table II). Note that all three groups of listeners received an SRM >4 dB on average, demonstrating the use of binaural cues to aid speech-in-noise perception in reverberant environments. The finding of an impact of reverberation on SRT in the spatially separated condition supported the hypothesis that speech-in-noise perception improved by partially restoring the distorted binaural cues through reducing RT from 1.1 to 0.4 s.

On the contrary, in the absence of spatial separation when target and maskers were co-located but of different

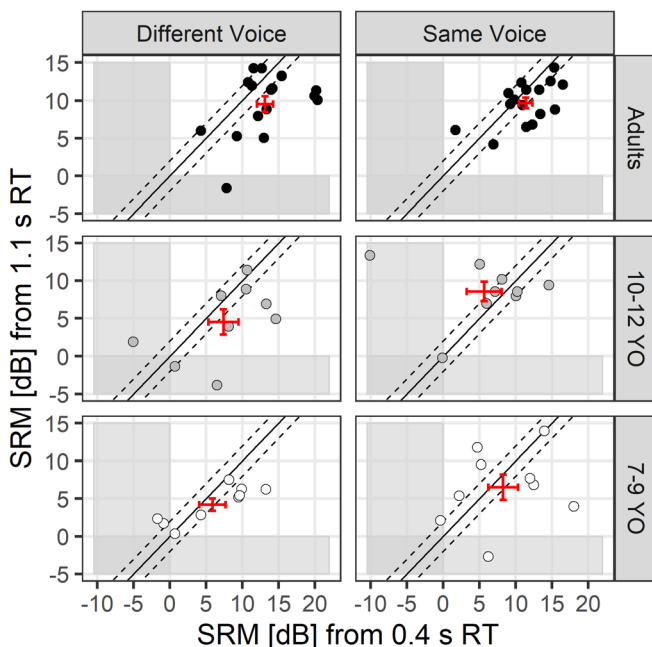


FIG. 4. (Color online) Individual data showing SRM from 0.4 s RT (x axis) and 1.1 s RT (y axis) separately for three age groups and for the same- versus different-voice conditions. Data falling on the solid diagonal lines (spanned by the dashed line at ± 2 dB) indicate similar SRMs for the two RTs. The shaded area on each panel indicates regions of negative SRM. Mean (± 1 SEM) SRMs are plotted in red error bars for both horizontal and vertical directions. One child in the 10–12 YO age group with an outlier value of SRM from 0.4 s was not included in the mean or SEM plotted.

voices (Fig. 2, upper left panel), the detrimental effect of reverberation on SRT weakened for all listener groups as compared to SRT increased over RT in other conditions. This finding suggested that children as young as 7–9 YO may already use voice cues effectively for speech-in-noise segregation, such that reducing RT, with enhanced temporal speech features, did not provide substantial release from masking. Interestingly, an intelligibility benefit from reducing reverberation was observed in the same-voice spatially co-located condition, but only among the 10–12 YO children and adults (Table I), suggesting that reduced temporal smearing was most effective for unmasking in the absence of other auditory cues (i.e., voice or spatial). However, the 7–9 YO children, who shared similar averaged SRT as the 10–12 YO children under 1.1 s RT, seemed unable to benefit from the enhanced temporal envelope by lowering RT to 0.4 s under the same-voice and spatially co-located conditions. While there is evidence to suggest sensory temporal resolution reaching maturity around 4–5 years old (Hall and Grose, 1994), our finding adds evidence to the idea that the protracted non-sensory maturation prolonged the development of using temporal cues (i.e., envelope cues) for speech-in-noise perception (Cabrera *et al.*, 2019).

It is worth noting that reverberation characteristics might contribute to unmasking. In the present study, the simulated reverberant environments had different T_{30} curves (Fig. 1), with a flat curve for the 0.4 s RT and a sloping curve for the 1.1 s RT as a function of octave band center frequencies. Specifically, the 1.1 s RT curve had an averaged T_{30} of 1.4 s between 125 and 250 Hz and 0.8 s above 4 kHz. It is not uncommon for realistic indoor spaces to have higher T_{30} in low-frequency regions, because most surface materials were less effective at absorbing low-frequency sounds than mid- to high-frequency sounds. One working hypothesis in current literature (Ljung *et al.*, 2009; Maccutcheon *et al.*, 2019) is the role of upward spread of masking in reverberant speech perception, in which slower energy decay in the low-frequency region (e.g., vowels) may exacerbate its masking of higher-frequency sounds (e.g., consonants). All three female talkers in this study had averaged first and second formants carrying the most important acoustic information for speech perception between 500 and 2000 Hz (see Table I). The T_{30} was rather consistent at 0.4 s and between 1 and 1.2 s for the two RT conditions suggesting similar energy decay within this frequency range; hence, the impact of upward spread of masking was considered minimal in both RT conditions. Furthermore, the largest T_{30} discrepancy was in the 125–250 Hz. For speech, this frequency region carries the most information on the fundamental frequency (see Table I) that is important for talker identification. With all three listener groups demonstrating sufficient use of the voice cues for stream segregation as previously discussed, it was unlikely that the 125–250 Hz low-frequency masking significantly impacted the mid- or high-frequency sounds either.

However, other speech cues that involve both spectral and temporal cues may be differentially influenced by the

amount of reverberant environment in each octave band. Previous work by Johnson (2000) suggested that, for children, consonant contrasts that varied in voicing, manner, and place of articulation were less salient in the presence of reverberation, particularly when it was coupled with noise (Johnson, 2000). While we cannot discern the differential impact of longer versus shorter RTs on such fine-grained acoustic cues in the present study, it is an area that is worthy of future investigation among school-age children as we found immaturity in their use of temporal cues.

With a significantly smaller SRM in both groups of children than adults, we found immaturity in stream segregation using binaural cues in reverberant environments even among the 10–12 YO children. This finding contradicted previous reports of earlier maturation when children were tested in a free-field setup (Misurelli and Litovsky 2012; 2015)². The results of this investigation are more consistent with studies indicating a longer developmental trajectory of SRM (Cameron *et al.*, 2011; Van Deun *et al.*, 2010; Vaillancourt *et al.*, 2008; Yuen and Yuan, 2014); children may be more vulnerable than adults to distortions in binaural cues from reverberation for unmasking in realistic complex auditory environments. Such vulnerability was also observed as more children experienced interference from distorted binaural cues when the symmetrical spatial separation was introduced. When we examined individual data of SRM, we observed four out of 18 children (but only one in 16 adults) to demonstrate negative SRMs, an interference or anti-benefit with increasing rather than decreasing SRTs by introducing the 90° target-masker separation, in one or more conditions.

Our findings on the role of talker similarity suggested that it might modulate the impact of reverberation on SRM. Note that the average SRM was similar under 1.1 s RT in both voice conditions. When the target and maskers were of different voices (Fig. 3), similar to everyday communication scenarios, we observed that adults achieved a generally larger SRM improvement with reducing RT than in the same-voice condition. This finding supported the alternative hypothesis in which spatial attention is impaired in reverberant environments and corroborated previous findings from adults (Oberem *et al.*, 2018; Ruggles *et al.*, 2011). Also, in the different-voice condition, the 10–12 YO children showed a similar general trend of SRM improvement as adults, even though the improvement did not reach statistical significance. In contrast, with voice cues removed in the same-voice conditions, such SRM improvement was completely absent for the older children group. Hence, we speculated that within the age range of 10–12 YO, some children might begin to draw benefit from reducing RT to improve SRM but only when the target and maskers had distinct voices. However, for the 7–9 YO children, the averaged SRM was similar across RTs in both voice cue conditions; reducing RT only provided an SRM improvement >1 dB for three out of nine children in this age group.

While adults clearly demonstrated SRM improvement with reducing RT, our findings suggested otherwise for

children. Until at least 12 years of age, children's SRM did not seem to benefit from reducing reverberation from 1.1 to 0.4 s RT. Even though at 0.4 s RT, a level within the ANSI recommendation for good classroom acoustics, the distortions to binaural cues from reverberation presented a substantial challenge for children that was similar to an environment with excessive reverberation at 1.1 s RT. When does SRM in reverberation reach maturity in typically developing children? Is there an ideal RT <0.4 s that better promotes SRM in young children? These are directions that warrant future investigations to guide our classroom acoustics designs for young children in helping them navigate complex auditory environments.

V. LIMITATIONS

In this study, we tested two relatively small groups of children. Even though a number of important effects on SRM, including reverberation, were detected, we did not observe the prolonged developmental trajectory into late adolescence on speech perception with two-talker babble as previously reported (Buss *et al.*, 2017). Overall, we used a combination of target sentence recognition with the two-talker, same-sex maskers, resulting in a difficult listening task with high informational masking even though the task better mimicked real-world listening. In conditions with the spatially co-located maskers, both children and adults had positive SRTs, suggesting the possibility of their relying on additional cues from target sentences at positive SNRs for release from informational masking (Swaminathan *et al.*, 2015). How such potential confound interacts with distortions on the speech envelopes from reverberation remains to be explored.

Our current investigation used a well-tested SRM paradigm implemented in virtual acoustic environments. But these virtual environments or experimental tasks did not perfectly emulate the everyday environments. Capturing the SRT thresholds at 50% was an efficient way to assess speech-in-noise performance, but it was limited in inferring speech comprehension experienced by children in their classroom settings. It is of interest for future work to consider how additional factors, such as visual input, from realistic classroom environments may influence children's spatial hearing abilities in reverberation. Furthermore, the size of SRM varies between studies depending on the speech materials used for the target and masker and is known to vary substantially across studies (Cameron *et al.*, 2011; Ching *et al.*, 2011; Corbin *et al.*, 2017; Van Deun *et al.*, 2010; Litovsky, 2005). Our findings of SRM in reverberation may be unique to the type of speech materials chosen for this investigation that aimed at maximizing informational masking, with the intention that children might rely more heavily on spatial cues for unmasking in the absence of other acoustic cues.

VI. SUMMARY AND CONCLUSION

In this investigation, we used state-of-the-art acoustic virtual reality technologies to examine the impact of

realistic reverberant environments on children's use of spatial hearing for speech-in-noise understanding. The major findings related to our research questions are listed below.

- (1) When reducing RT from 1.1 to 0.4 s, 7–12 YO children demonstrated SRT improvement in the condition when the target and maskers were of different voices and spatially separated. Reducing distortions in the binaural cues improved intelligibility for children.
- (2) We found immaturity in SRM in reverberant environments until at least 10–12 years old. Previous studies that measured SRM in anechoic environments may have overestimated children's ability to use binaural cues for stream segregation.
- (3) Reducing RT from 1.1 to 0.4 s did not improve SRM for either group of children, although the 10–12 YOs showed a general improvement trend when the target and maskers were of different voices. Only adults experienced more SRM at the shorter RT.
- (4) Talker similarity between the target and masker may modulate the effect of reverberation on SRM only among adults, but not for the children in this investigation.

In conclusion, 7–12 YO children's masked speech perception improved when reducing reverberation from 1.1 to 0.4 s in the presence of spatial separation between the target and two-talker maskers. Children demonstrated 3–8 dB SRM on average in reverberant environments, but this benefit was much smaller than that observed for adults (10–13 dB). The slight distortion from reverberation may be so disruptive for children to access binaural cues, particularly the younger ones in the 7–9 YO age range, that they were unable to draw benefits from reducing RT to gain a larger SRM.

ACKNOWLEDGMENTS

We would like to acknowledge funding support from the European Union's Seventh Framework Programme: Improving Children's Auditory Rehabilitation (iCARE, Grant No. ITN FP7-607139), the Excellence Initiative of the German federal and state governments (ERS Boost Fund 2014; Grant No. OPBF090) awarded to JF. ZEP received support from the National Institutes of Health (NIH) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant Nos. T32HD007489 and U54HD090256) that was awarded to the Waisman Center at the University of Wisconsin-Madison.

We are grateful for all the children and their families as well as adults from the local Aachen area who participated in this study. We would also like to thank Lukas Aspöck and Ramona Bomhardt for their technical inputs during the development of this study and Karin Loh, Alokparna Ray, Arndt Brandl, and Laura Bell for assistance during material preparation and data collection.

¹See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0006752> for method to estimate individual psychometric functions for speech reception threshold and for method to estimate individual variabilities.

²SRM was calculated based on SRTs at 80% word accuracy in Misurelli and Litovsky (2012; 2015), but on SRTs at 50% keyword accuracy in the present study. There was potential difference in SRM estimated at a higher percentage correct. To validate our comparison with the Misurelli and Litovsky studies, we performed follow-up analysis to assess the difference in SRM if estimated at 70% correct. The psignifit curve estimation did not always generate reliable SRT estimated at 80% correct for some participants but had reliable SRTs at 70% correct for all participants. We saw significantly smaller SRM at 70% than SRM at 50% correct by 1.4 dB ($p < 0.001$) among adults, but non-significant differences for the 10–12 YO group (smaller SRM70 by 0.3 dB, $p = 0.42$) and for the 7–9 YO group (smaller SRM70 by 0.3 dB, $p = 0.06$). All SRM changes by estimating at a higher percentage correct SRTs are less than the error margin of 2 dB (i.e., smallest step size). Hence, we concluded that SRMs at 50% correct from the present study could be used to compare with those in Misurelli and Litovsky (2012; 2015).

Acoustical Society of America (2002). "ANSI/ASA S12.60-2002 (R2009) American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools," Stand. Secr.

Beutelmann, R., and Brand, T. (2006). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 331–342.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Blauert, J., and Lindemann, W. (1986a). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *J. Acoust. Soc. Am.* **79**, 806–813.

Blauert, J., and Lindemann, W. (1986b). "Auditory spaciousness: Some further psychoacoustic analyses," *J. Acoust. Soc. Am.* **80**, 533–542.

Boersma, P., and Weenink, D. (2021). "Praat: doing phonetics by computer," <http://www.praat.org/> (Last viewed August 30, 2020).

Bomhardt, R., and Fels, J. (2014). "Analytical interaural time difference model for the individualization of arbitrary head-related impulse responses," in *Audio Engineering Society Convention 137*, Audio Engineering Society.

Bomhardt, R., Lins, M., and Fels, J. (2016). "Analytical ellipsoidal model of interaural time differences for the individualization of head-related impulse responses," *J. Audio Eng. Soc.* **64**, 882–894.

Bomhardt, R., Patiño Mejía, I. C., Zell, A., and Fels, J. (2018). "Required measurement accuracy of head dimensions for modeling the interaural time difference," *J. Audio Eng. Soc.* **66**, 114.

Brown, A. D., Stecker, G. C., and Tollin, D. J. (2015). "The precedence effect in sound localization," *J. Assoc. Res. Otolaryngol.* **16**, 1–28.

Buss, E., Leibold, L. J., Porter, H. L., and Grose, J. H. (2017). "Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing," *J. Acoust. Soc. Am.* **141**, 2650–2660.

Cabrera, L., Varnet, L., Buss, E., Rosen, S., and Lorenzi, C. (2019). "Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity," *J. Acoust. Soc. Am.* **146**, 2415.

Cameron, S., and Dillon, H. (2007). "Development of the listening in spatialized noise-sentences test (LISN-S)," *Ear Hear.* **28**, 196–211.

Cameron, S., Glyde, H., and Dillon, H. (2011). "Listening in spatialized noise—sentences test (LiSN-S): Normative and retest reliability data for adolescents and adults up to 60 years of age," *J. Am. Acad. Audiol.* **22**, 697–709.

Ching, T. Y. C., van Wanrooy, E., Dillon, H., and Carter, L. (2011). "Spatial release from masking in normal-hearing children and children who use hearing aids," *J. Acoust. Soc. Am.* **129**, 368–375.

Clifton, R. K., Gwiazda, J., Bauer, J. A., Clarkson, M. G., and Held, R. M. (1988). "Growth in head size during infancy: Implications for sound localization," *Dev. Psychol.* **24**, 477–483.

Corbin, N. E., Buss, E., and Leibold, L. J. (2017). "Spatial release from masking in children: Effects of simulated unilateral hearing loss," *Ear Hear.* **38**, 223–235.

Deroche, M. L. D., Culling, J. F., Lavandier, M., and Gracco, V. L. (2017). "Reverberation limits the release from informational masking obtained in the harmonic and binaural domains," *Atten. Percept. Psychophys.* **79**, 363–379.

Fels, J. (2008). "From children to adults: How binaural cues and ear canal impedances grow," *J. Acoust. Soc. Am.* **124**(6), 3359.

Fels, J., Buthmann, P., and Vorländer, M. (2004). "Head-related transfer functions of children," *Acta Acust. united Acust.* **90**(5), 918–927.

Fels, J., and Vorländer, M. (2009). "Anthropometric parameters influencing head-related transfer functions," *Acta Acust. united Acust.* **95**(2), 331–342.

Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R* (Sage Publications, London, UK).

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.* **106**, 3578–3588.

Frund, I., Haenel, N. V., and Wichmann, F. A. (2011). "Inference for psychometric functions in the presence of nonstationary behavior," *J. Vis.* **11**, 16.

Goupell, M. J., and Hartmann, W. M. (2006). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," *J. Acoust. Soc. Am.* **119**, 3971–3986.

Goupell, M. J., Yu, G., and Litovsky, R. Y. (2012). "The effect of an additional reflection in a precedence effect experiment," *J. Acoust. Soc. Am.* **131**, 2958–2967.

Griffin, A. M., Poissant, S. F., and Freyman, R. L. (2019). "Speech-in-noise and quality-of-life measures in school-aged children with normal hearing and with unilateral hearing loss," *Ear Hear.* **40**, 887–904.

Hall, J. W., and Grose, J. H. (1994). "Development of temporal resolution in children as measured by the temporal modulation transfer function," *J. Acoust. Soc. Am.* **96**, 150–154.

Hartmann, W. M. (1983). "Localization of sound in rooms," *J. Acoust. Soc. Am.* **74**, 1380–1391.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.* **115**, 833–843.

Hidaka, T., Beranek, L. L., and Okano, T. (1995). "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *J. Acoust. Soc. Am.* **98**, 988–1007.

Hochmair-Desoyer, I., Schulz, E., Moser, L., and Schmidt, M. (1997). "The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users," *Am. J. Otol.* **18**(6 Suppl.), S83–S83.

Houtgast, T., and Steeneken, H. J. M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**, 1069.

Johnson, C. E. (2000). "Children's phoneme identification in reverberation and noise," *J. Speech Lang. Hear. Res.* **43**, 144–157.

Johnson, D., and Lee, H. (2019). "Perceptual threshold of apparent source width in relation to the azimuth of a single reflection," *J. Acoust. Soc. Am.* **145**, EL272–EL276.

Johnstone, P. M., and Litovsky, R. Y. (2006). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," *J. Acoust. Soc. Am.* **120**, 2177–2189.

Kidd, G., Mason, C. R., Brughera, A., and Hartmann, W. M. (2005). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," *Acta Acust. united Acust.* **91**, 526–536.

Klatte, M., Lachmann, T., and Meis, M. (2010). "Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting," *Noise Heal.* **12**, 270.

Kuttruff, H. (2009). *Room Acoustics*, 5th ed. (CRC Press, London, UK).

Leibold, L. J., and Buss, E. (2013). "Children's identification of consonants in a speech-shaped noise or a two-talker masker," *J. Speech Lang. Hear. Res.* **56**, 1144–1155.

Leibold, L. J., and Buss, E. (2019). "Masked speech recognition in school-age children," *Front. Psychol.*

Leibold, L. J., and Neff, D. L. (2007). "Effects of masker-spectral variability and masker fringes in children and adults," *J. Acoust. Soc. Am.* **121**, 3666.

Leibold, L. J., and Neff, D. L. (2011). "Masking by a remote-frequency noise band in children and adults," *Ear Hear.* **32**, 663–666.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.

Litovsky, R. Y. (1997). "Developmental changes in the precedence effect: Estimates of minimum audible angle," *J. Acoust. Soc. Am.* **102**, 1739–1745.

- Litovsky, R. Y. (2005). "Speech intelligibility and spatial release from masking in young children," *J. Acoust. Soc. Am.* **117**, 3091–3099.
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., Guzman, S. J., Litovsky, R. Y., and Colburn, H. S. (1999). "The precedence effect," *J. Acoust. Soc. Am.* **106**, 1633–1654.
- Litovsky, R. Y., and Godar, S. P. (2010). "Difference in precedence effect between children and adults signifies development of sound localization abilities in complex listening tasks," *J. Acoust. Soc. Am.* **128**, 1979–1991.
- Ljung, R., Sörqvist, P., Kjellberg, A., and Green, A.-M. (2009). "Poor listening conditions impair memory for intelligible lectures: Implications for acoustic classroom standards," *Build. Acoust.* **16**, 257–265.
- Maccutcheon, D., Pausch, F., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., Fels, J., and Ljung, R. (2019). "The contribution of individual differences in memory span and language ability to spatial release from masking in young children," *J. Speech Lang. Hear. Res.* **62**, 3741–3751.
- Marrone, N., Mason, C. R., and Kidd, G. (2008). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," *J. Acoust. Soc. Am.* **124**, 3064–3075.
- Masiero, B., and Vorländer, M. (2014). "A framework for the calculation of dynamic crosstalk cancellation filters," *IEEE Trans. Audio Speech Lang. Process.* **22**, 1345–1354.
- McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D., and Brennan, M. (2019). "Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss," *Front. Neurosci.* **13**, 1093.
- Middlebrooks, J. C. (1999). "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.* **106**, 1480.
- Misurelli, S. M., and Litovsky, R. Y. (2012). "Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry," *J. Acoust. Soc. Am.* **132**, 380–391.
- Misurelli, S. M., and Litovsky, R. Y. (2015). "Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity," *J. Acoust. Soc. Am.* **138**, 319–331.
- Muñoz, R. V., Aspöck, L., and Fels, J. (2019). "Spatial release from masking under different reverberant conditions in young and elderly subjects: Effect of moving or stationary maskers at circular and radial conditions," *J. Speech Lang. Hear. Res.* **62**, 3582–3595.
- Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (1989). "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.* **86**, 1259–1265.
- Nábělek, A. K., and Robinson, P. K. (1982). "Monaural and binaural speech perception in reverberation for listeners of various ages," *J. Acoust. Soc. Am.* **71**, 1242–1248.
- Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (2010). "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," *Ear Hear.* **31**, 336–344.
- Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (2012). "Measuring speech recognition in children with cochlear implants in a virtual classroom," *J. Speech Lang. Hear. Res.* **55**, 532–540.
- Oberem, J., Seibold, J., Koch, I., and Fels, J. (2018). "Intentional switching in auditory selective attention: Exploring attention shifts with different reverberation times," *Hear. Res.* **359**, 32–39.
- Pausch, F., Aspöck, L., Vorländer, M., and Fels, J. (2018). "An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss," *Trends Hear.* **22**, 233121651880087.
- Pausch, F., and Fels, J. (2020). "Localization performance in a binaural real-time auralization system extended to research hearing aids," *Trends Hear.* **24**, 233121652090870.
- Pelzer, S., Aspöck, L., Schröder, D., and Vorländer, M. (2014). "Interactive real-time simulation and auralization for modifiable rooms," *Build. Acoust.* **21**, 65.
- Peng, Z. E., and Wang, L. M. (2016). "Effects of noise, reverberation and foreign accent on native and non-native listeners' performance of English speech comprehension," *J. Acoust. Soc. Am.* **139**, 2772–2783.
- Prodi, N., Visentin, C., and Feletti, A. (2013). "On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence," *J. Acoust. Soc. Am.* **133**, 255–268.
- Rakerd, B., and Hartmann, W. M. (1985). "Localization of sound in rooms II: The effects of a single reflecting surface," *J. Acoust. Soc. Am.* **78**, 524–533.
- Rakerd, B., Hartmann, W. M., and Rakerd, B. (2010). "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise," *J. Acoust. Soc. Am.* **128**, 3052–3063.
- Rennies, J., Brand, T., Kollmeier, B., and Brand, T. (2011). "Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet," *J. Acoust. Soc. Am.* **130**, 2999–3012.
- Rennies, J., and Kidd, G. (2018). "Benefit of binaural listening as revealed by speech intelligibility and listening effort," *J. Acoust. Soc. Am.* **144**, 2147–2159.
- Robinson, P. W., Pätynen, J., Lokki, T., Suk Jang, H., Yong Jeon, J., and Xiang, N. (2013). "The role of diffusive architectural surfaces on auditory spatial discrimination in performance venues," *J. Acoust. Soc. Am.* **133**, 3940–3950.
- Ruggles, D., Shinn-Cunningham, B., and Unningham, B. A. S. H. (2011). "Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners," *J. Assoc. Res. Otolaryngol.* **12**, 395–405.
- Schimmel, S. M., Müller, M. F., and Dillier, N. (2009). "A fast and accurate 'shoebox' room acoustics simulator," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, New York), pp. 241–244.
- Schmidt, A. (1995). "A new digital artificial head measurement system," *Acta Acust. united Acust.* **81**, 416–520.
- Schröder, D., and Vorländer, M. (2011). "RAVEN: A real-time framework for the auralization of interactive virtual environments," in *Proceedings of Forum Acusticum*, Aalborg, Denmark, pp. 1541–1546.
- Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (2005). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," *J. Acoust. Soc. Am.* **117**, 3100–3115.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., and Patel, A. D. (2015). "Musical training, individual differences and the cocktail party problem," *Sci. Rep.* **5**, 11628.
- Vaillancourt, V., Laroche, C., Giguère, C., and Soli, S. D. (2008). "Establishment of age-specific normative data for the Canadian French version of the hearing in noise test for children," *Ear Hear.* **29**, 453–466.
- Valente, D. L., Plevinsky, H. M., Franco, J. M., Heinrichs-Graham, E. C., and Lewis, D. E. (2012). "Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children," *J. Acoust. Soc. Am.* **131**, 232–246.
- Van Deun, L., Van Wieringen, A., and Wouters, J. (2010). "Spatial speech perception benefits in young children with normal hearing and cochlear implants," *Ear Hear.* **31**(5), 702–713.
- Wefers, F. (2015). "Partitioned convolution algorithms for real-time auralization," Ph.D. dissertation, RWTH Aachen Univ.
- Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2013). "Measuring the apparent width of auditory sources in normal and impaired hearing," *Adv. Exp. Med. Biol.* **787**, 303–310.
- Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2014). "The perception of apparent auditory source width in hearing-impaired adults," *J. Acoust. Soc. Am.* **135**, 3548–3559.
- Wightman, F. L., Kistler, D. J., and O'Bryan, A. (2010). "Individual differences and age effects in a dichotic informational masking paradigm," *J. Acoust. Soc. Am.* **128**, 270–279.
- Wroblewski, M., Lewis, D. E., Valente, D. L., and Stelmachowicz, P. G. (2012). "Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults," *Ear Hear.* **33**, 731–744.
- Yang, W., and Bradley, J. S. (2009). "Effects of room acoustics on the intelligibility of speech in classrooms for young children," *J. Acoust. Soc. Am.* **125**, 922–933.
- Yuen, K. C. P., and Yuan, M. (2014). "Development of spatial release from masking in mandarin-speaking children with normal hearing," *J. Speech Lang. Hear. Res.* **57**, 2005–2023.